

# On The Exact Recovery Condition of Simultaneous Orthogonal Matching Pursuit

Jean-François Determe\* Jérôme Louveaux† Laurent Jacques† François Horlin\*

December 16, 2015

## Abstract

Several exact recovery criteria (ERC) ensuring that orthogonal matching pursuit (OMP) identifies the correct support of sparse signals have been developed in the last few years. These ERC rely on the restricted isometry property (RIP), the associated restricted isometry constant (RIC) and sometimes the restricted orthogonality constant (ROC). In this paper, three of the most recent ERC for OMP are examined. The contribution is to show that these ERC remain valid for a generalization of OMP, entitled simultaneous orthogonal matching pursuit (SOMP), that is capable to process several measurement vectors simultaneously and return a common support estimate for the underlying sparse vectors. The sharpness of the bounds is also briefly discussed in light of previous works focusing on OMP.

## 1 Introduction

Recovering a high dimensional sparse signal by acquiring it through a linear measurement process returning fewer observations than its dimension is a problem often encountered in the digital signal processing literature. The field of research associated with such problems is often known to researchers as *compressed sensing* or *compressive sensing* (CS) [12].

We define the support of a vector  $\mathbf{x} \in \mathbb{R}^n$  as  $\text{supp}(\mathbf{x}) := \{j \in [n] : x_j \neq 0\}$  where  $[n]$  denotes the set  $\{1, 2, \dots, n\}$  and  $x_j$  denotes the  $j$ th entry of  $\mathbf{x}$ . A vector is said to be  $s$ -sparse whenever its support exhibits a cardinality equal to or lower than  $s$ .

### 1.1 Signal model

In this paper, we focus on a framework involving

1.  $K$  sparse signals  $\mathbf{x}_k \in \mathbb{R}^n$  to be recovered ( $1 \leq k \leq K$ ),
2. a common linear measurement process described by the matrix  $\Phi \in \mathbb{R}^{m \times n}$ ,
3.  $K$  measurement vectors  $\mathbf{y}_k \in \mathbb{R}^m$  gathering the observations of each sparse signal when acquired through  $\Phi$ :  $\mathbf{y}_k = \Phi \mathbf{x}_k$ .

---

\*Jean-François Determe and François Horlin are with the OPERA Wireless Communications Group, Université Libre de Bruxelles, 1050 Brussels, Belgium. E-mail: jdeterme@ulb.ac.be, fhorlin@ulb.ac.be. Jean-François Determe is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

†Laurent Jacques and Jérôme Louveaux are with the ICTEAM département, Université Catholique de Louvain. E-mail: laurent.jacques@uclouvain.be, jerome.louveaux@uclouvain.be. Laurent Jacques is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

To simplify the signal model, we introduce Equation (1) to summarize the  $K$  equations  $\mathbf{y}_k = \Phi \mathbf{x}_k$  into a single one:

$$\mathbf{Y} = \Phi \mathbf{X} \quad (1)$$

where  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_K) \in \mathbb{R}^{m \times K}$  and  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_K) \in \mathbb{R}^{n \times K}$ . Using this formulation, the support of  $\mathbf{X}$ , denoted by  $\text{supp}(\mathbf{X})$ , is equal to the joint support  $S := \cup_{k \in [K]} \text{supp}(\mathbf{x}_k)$ .

When a model involves one measurement vector, it is referred to as a single measurement vector (SMV) model while models incorporating  $K > 1$  measurement vectors are multiple measurement vector (MMV) models [14].

The columns of  $\Phi$  are often referred to as the *atoms*. This terminology being typically associated with dictionaries, it is worth emphasizing that the problem of recovering a  $s$ -sparse vector  $\mathbf{x}$  on the basis of the measurement vector  $\mathbf{y} = \Phi \mathbf{x}$  is equivalent to finding  $s$  columns (or atoms) of the (dictionary) matrix  $\Phi$  that fully express  $\mathbf{y}$  when using the proper linear combination. The notion of atom will thus be used in the rest of this paper as it simplifies the mathematical discussions that follow.

We now introduce additional notions that are used afterwards. For  $0 < p < \infty$  and  $\mathbf{x} \in \mathbb{R}^n$ , we define the norms  $\|\mathbf{x}\|_p := (\sum_{j=1}^n |x_j|^p)^{1/p}$  and  $\|\mathbf{x}\|_\infty := \max_{j \in [n]} |x_j|$ . In this paper, every vector should be considered as a column vector. Also, for  $S \subseteq [n]$ , the quantity  $\mathbf{x}_S$  denotes the vector formed by the entries of  $\mathbf{x}$  indexed by  $S$ . Similarly, for a matrix  $\Phi \in \mathbb{R}^{m \times n}$ , we define  $\Phi_S$  as the matrix formed by the columns of  $\Phi$  indexed within  $S$ . The Moore-Penrose pseudoinverse of any matrix  $\Phi$  is denoted by  $\Phi^+$  and its transpose is given by  $\Phi^T$ . Finally, the inner product of two vectors  $\mathbf{x}$  and  $\mathbf{y}$  is written as  $\langle \mathbf{x}, \mathbf{y} \rangle$  and is equal to  $\mathbf{x}^T \mathbf{y}$ .

## 1.2 Simultaneous orthogonal matching pursuit

Several algorithms exhibiting varying computational complexities have been investigated to address the problem above. For the SMV case, the greedy algorithm entitled orthogonal matching pursuit (OMP) [9, 20] is a classical choice because its complexity is lower than that of other algorithms such as  $\ell_1$ -minimization [13].

If the  $K$  sparse signals  $\mathbf{x}_k$  possess similar supports, *i.e.*, their joint support  $S := \cup_{k \in [K]} \text{supp}(\mathbf{x}_k)$  possesses a cardinality that is comparable to those of the individual supports  $\text{supp}(\mathbf{x}_k)$ , then it is interesting to perform a joint and common estimation of their supports [10, 16]. The simultaneous orthogonal matching pursuit (SOMP) algorithm [22], which is described in Algorithm 1, is an extension of OMP to the MMV case and performs a joint support recovery.

As shown in Algorithm 1, at each iteration  $t$ , SOMP adds to the estimated support the index  $j_t$  of the atom  $\phi_{j_t}$  maximizing the metric  $\|(\mathbf{R}^{(t)})^T \phi_j\|_1 = \sum_{k=1}^K |\langle \phi_j, \mathbf{r}_k^{(t)} \rangle|$  (steps 4 and 5) where  $\mathbf{r}_k^{(t)}$  denotes the  $k$ th column of the residual matrix  $\mathbf{R}^{(t)}$ . Each measurement vector  $\mathbf{y}_k$  is then projected onto the orthogonal complement of  $\text{span}(\Phi_{S_{t+1}})$ , denoted by  $\text{span}(\Phi_{S_{t+1}})^\perp$ , during steps 6 and 7. The algorithm terminates when the prescribed number of iterations  $s$  has been reached. It is worth noticing that an atom cannot be picked twice as, once chosen, the projection onto  $\text{span}(\Phi_{S_{t+1}})^\perp$  ensures that  $\langle \phi, \mathbf{r}_k^{(t+1)} \rangle = 0$  if  $\phi \in S_t$ .

## 1.3 Definitions

We define the concepts needed to state the results of Section 2. First of all, the matrix  $\Phi$  is said to satisfy the restricted isometry property (RIP) [4] of order  $s$  with restricted isometry constant

ALGORITHM 1:  
Simultaneous orthogonal matching pursuit (SOMP)

**Require:**  $\mathbf{Y} \in \mathbb{R}^{m \times K}$ ,  $\Phi \in \mathbb{R}^{m \times n}$ ,  $s \geq 1$   
1: Initialization:  $\mathbf{R}^{(0)} \leftarrow \mathbf{Y}$  and  $S_0 \leftarrow \emptyset$   
2:  $t \leftarrow 0$   
3: **while**  $t < s$  **do**  
4:   Determine the atom of  $\Phi$  to be included in the support:  
 $j_t \leftarrow \operatorname{argmax}_{j \in [n]} (\|(\mathbf{R}^{(t)})^T \phi_j\|_1)$   
5:   Update the support :  $S_{t+1} \leftarrow S_t \cup \{j_t\}$   
6:   Projection of each measurement vector onto  $\operatorname{span}(\Phi_{S_{t+1}})$ :  
 $\mathbf{Y}^{(t+1)} \leftarrow \Phi_{S_{t+1}}^+ \Phi_{S_{t+1}}^T \mathbf{Y}$   
7:   Projection of each measurement vector onto  $\operatorname{span}(\Phi_{S_{t+1}})^\perp$  :  
 $\mathbf{R}^{(t+1)} \leftarrow \mathbf{Y} - \mathbf{Y}^{(t+1)}$   
8:    $t \leftarrow t + 1$   
9: **end while**  
10: **return**  $S_s$  {Support at last step}

(RIC)  $\delta_s$  (of order  $s$ ) whenever

$$(1 - \delta_s) \|\mathbf{u}\|_2^2 \leq \|\Phi \mathbf{u}\|_2^2 \leq (1 + \delta_s) \|\mathbf{u}\|_2^2 \quad (2)$$

holds for all  $s$ -sparse vectors  $\mathbf{u}$ . Thus, the RIP ensures that the linear operator  $\Phi$  maintains the  $\ell_2$ -norm of  $s$ -sparse signals up to a certain extent that is quantified by means of the RIC  $\delta_s$ . Furthermore, if  $\mathbf{u}$  is supported onto  $S$ , the quantity  $\|\Phi \mathbf{u}\|_2^2$  is equal to  $\|\Phi_S \mathbf{u}_S\|_2^2 = \mathbf{u}_S^T (\Phi_S^T \Phi_S) \mathbf{u}_S$ . The RIP therefore ensures that  $1 - \delta_s \leq \lambda_{\min}(\Phi_S^T \Phi_S) \leq \lambda_{\max}(\Phi_S^T \Phi_S) \leq 1 + \delta_s$  for all the supports  $S$  of cardinality equal to or lower than  $s$  where  $\lambda_{\min}$  and  $\lambda_{\max}$  denote the minimal and maximal eigenvalues, respectively. Also, it is easy to show that  $\delta_s \leq \delta_{s+1}$ .

The  $(\alpha, \alpha')$ -restricted orthogonality constant (ROC) [2] is defined as the smallest real number  $\theta_{\alpha, \alpha'}$  for which

$$|\langle \Phi \mathbf{c}, \Phi \mathbf{c}' \rangle| \leq \theta_{\alpha, \alpha'} \|\mathbf{c}\|_2 \|\mathbf{c}'\|_2 \quad (3)$$

holds for every  $\mathbf{c}, \mathbf{c}' \in \mathbb{R}^m$  exhibiting disjoint supports of cardinality  $\alpha$  and  $\alpha'$ , respectively. Thus, the ROC quantifies how vectors with disjoint supports stay approximately orthogonal after projection by  $\Phi$ .

The ROC and the RIP are linked by the inequality [3, Lemma 2.1]  $\theta_{\alpha, \alpha'} \leq \delta_{\alpha+\alpha'}$  which indicates that the RIC can play a role similar to that of the ROC, albeit in a less sharp manner. Another similar inequality has been obtained in [23, Section 2.3] and is given by  $\theta_{1, \alpha'} \leq \sqrt{\alpha' / (\alpha' - 1)} \delta_{\alpha'}$  whenever  $\alpha' \geq 2$ . Another upper bound of  $\theta_{1, \alpha'}$  has been recently obtained in [28, Lemma II.3] where the so-called 2-coherence of the dictionary matrix, denoted by  $\nu_{\alpha'}$ , is used. They have shown [28, Lemma II.2] that  $\nu_{\alpha'} \leq \delta_{\alpha'+1}$  so that the inequality  $\theta_{1, \alpha'} \leq \nu_{\alpha'}$  is sharper than  $\theta_{1, \alpha'} \leq \delta_{1+\alpha'}$ . The developments presented hereafter use the RIC-based inequality so that only the RIC intervenes in the final results. However, expressing our results using  $\nu_{\alpha'}$  instead of  $\delta_{1+\alpha'}$  is straightforward.

Finally, it is worth defining the  $\ell_\infty$ -induced norm for matrices as  $\|\Phi\|_{\infty \rightarrow \infty} := \sup_{\|\phi\|_\infty=1} \|\Phi \phi\|_\infty$  (where  $\Phi \in \mathbb{R}^{m \times n}$ ) that can be computed as  $\|\Phi\|_{\infty \rightarrow \infty} = \max_{i \in [m]} \sum_{j=1}^n |\phi_{i,j}|$  [15, Lemma A.5]. This quantity is interesting as it allows to write, for  $A \subseteq [n]$ ,  $\max_{j \in A} (\|(\mathbf{R}^{(t)})^T \phi_j\|_1) = \|\Phi_A^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$ , which is reminiscent of the decision metric of SOMP. Some authors choose to

write the  $\ell_\infty$ -induced norm of  $\Phi$  as  $\|\Phi\|_\infty$  but, to avoid confusions, we prefer to emphasize the distinction between the  $\ell_\infty$ -norms for vectors and matrices as both coexist in Section 4.

## 2 Contribution and related work

The main contribution of this paper is to extend a recent exact recovery criterion (ERC) for OMP to its MMV counterpart, *i.e.*, SOMP. An ERC is a sufficient condition to ensure that the algorithm commits no mistake. The cornerstone of the results presented in this paper is given by Lemma 1.

**Lemma 1** (A RIP and ROC-based lower bound on the maximal residual projection). *Let  $\mathbf{X} \in \mathbb{R}^{n \times K}$  possess the support  $S$ . Let  $\Phi \in \mathbb{R}^{m \times n}$  admit the RIC  $\delta_{|S|} < 1$  and the  $(1, |S|)$ -ROC  $\theta_{1,|S|} < 1$ . Furthermore,  $\mathbf{P}^{(t)} = \Phi_{S_t} \Phi_{S_t}^+$  denotes the orthogonal projector onto  $\text{span}(\Phi_{S_t})$  where  $S_t \subseteq S$ , *i.e.*, only correct atoms have been included to the estimated support before iteration  $t$ . Let  $\mathbf{R}^{(t)}$  be equal to  $(\mathbf{I} - \mathbf{P}^{(t)})\mathbf{Y} = (\mathbf{I} - \mathbf{P}^{(t)})\Phi\mathbf{X}$ . Then,*

$$\frac{\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}}{\|\Phi_{\bar{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}} \geq \frac{1 - \delta_{|S|}}{\theta_{1,|S|} \sqrt{|S|}} \quad (4)$$

where  $\bar{S}$  is the relative complement of  $S$  with respect to  $[n]$ .

Lemma 1 establishes a lower bound on the ratio of the SOMP metric obtained for the correct atoms to that obtained for the incorrect ones. In that sense, and as it will be clarified in Theorem 1, it straightforwardly provides an ERC guaranteeing that SOMP commits no error when picking atoms. We now propose a corollary of Lemma 1 that only relies on the RIC.

**Lemma 2** (RIP lower bounds on the maximal residual projection). *Let  $\mathbf{X} \in \mathbb{R}^{n \times K}$  possess the support  $S$ . Let  $\Phi \in \mathbb{R}^{m \times n}$  admit the RIC  $\delta_{|S|} < 1$ . Furthermore,  $\mathbf{P}^{(t)} = \Phi_{S_t} \Phi_{S_t}^+$  denotes the orthogonal projector onto  $\text{span}(\Phi_{S_t})$  where  $S_t \subseteq S$ , *i.e.*, only correct atoms have been included to the estimated support before iteration  $t$ . Let  $\mathbf{R}^{(t)}$  be equal to  $(\mathbf{I} - \mathbf{P}^{(t)})\mathbf{Y} = (\mathbf{I} - \mathbf{P}^{(t)})\Phi\mathbf{X}$ . Then, both inequalities below hold*

$$\frac{\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}}{\|\Phi_{\bar{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}} \geq \frac{1 - \delta_{|S|+1}}{\delta_{|S|+1} \sqrt{|S|}} \quad (5)$$

$$\frac{\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}}{\|\Phi_{\bar{S}}^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}} \geq \frac{(1 - \delta_{|S|}) \sqrt{|S| - 1}}{\delta_{|S|} |S|} \quad (6)$$

where  $\bar{S}$  is the relative complement of  $S$  with respect to  $[n]$ .

Compared to former works that directly derived the ERC [7, 19, 23–25], we believe that Lemma 1 and Lemma 2 are interesting as they quantify the robustness of the decisions made at each iteration of SOMP in the noiseless case. Such quantities can then be used to produce theoretical analyses of greedy algorithms in a noisy setting (see [1, 7, 10]). Other similar works analysing OMP in a noisy environment include [21] and [28]. The analysis presented in [7] actually uses a result fundamentally identical to Lemma 1 for  $K = 1$  to conduct a theoretical analysis of OMP inspired from [1]. We now state the three ERC deriving from Lemma 1 and Lemma 2.

**Theorem 1** (Several RIP and ROC-based ERC for SOMP). *Let  $\mathbf{X} \in \mathbb{R}^{n \times K}$  possess the support  $S$ . Let  $\Phi \in \mathbb{R}^{m \times n}$  admit the RIC  $\delta_{|S|} < 1$  and the  $(1, |S|)$ -ROC  $\theta_{1,|S|} < 1$ . Then, SOMP commits no error and identifies the full support of  $\mathbf{X}$  at the end of iteration  $|S| - 1$  whenever at least one of the three conditions below hold:*

$$\frac{1 - \delta_{|S|}}{\theta_{1,|S|}\sqrt{|S|}} > 1 \quad (\text{ERC1})$$

$$\delta_{|S|+1} < \frac{1}{\sqrt{|S|} + 1} \quad (\text{ERC2})$$

$$(\text{for } |S| \geq 2) \quad \delta_{|S|} < \frac{\sqrt{|S|} - 1}{\sqrt{|S|} - 1 + |S|}. \quad (\text{ERC3})$$

As demonstrated in Section 4.3, Theorem 1 is a straightforward consequence of Lemma 1 and Lemma 2. The authors of [24] and [7] independently obtained (ERC1) for OMP. To the best of the authors' knowledge, the second ERC was first obtained simultaneously in [19] and [25] while (ERC3) was initially published in [23], both ERC being derived for OMP.

Regarding older works, it is also worth pointing out that the ERC  $\delta_{|S|+1} < 1/((1+\sqrt{2})\sqrt{|S|})$ , first obtained in [17, Theorem 5.2] for OMP, has been shown to remain valid for SOMP in [11, Corollary 1]. Thereby, the authors of [11] also proved that the older ERC  $\delta_{|S|+1} < 1/(3\sqrt{|S|})$ , initially derived in [8, Theorem 3.1] for OMP, remains correct for SOMP as  $\delta_{|S|+1} < 1/(3\sqrt{|S|})$  is implied by  $\delta_{|S|+1} < 1/((1+\sqrt{2})\sqrt{|S|})$ . Very recently, (ERC2) was extended to SOMP in [27, Remark 1]. However, the extension to SOMP of both (ERC1) and (ERC3) is a novel result. In [18], the author has derived the ERC  $\delta_{|S|+1} < 1/\sqrt{|S|} + 1$ , which is sharper than (ERC2). Combining the ideas developed in [18] and our paper could possibly extend this ERC to SOMP.

Finally, we would like to point out that, if any of the considered ERC holds, running  $K$  independent executions of OMP instead of a single instance of SOMP would enable one to retrieve the individual supports  $\text{supp}(\mathbf{x}_k)$  ( $1 \leq k \leq K$ ) and, by extension, the joint support  $S$ . While it may seem to undermine the interest of this work, the following observations convince otherwise:

1. If one of the considered ERC guarantees that each one of the  $K$  instances of OMP returns the correct support of each sparse vector  $\mathbf{x}_k$ , then SOMP is also guaranteed to return the correct joint support so that there is no penalty switching from OMP to SOMP, except maybe that SOMP returns a joint support instead of possibly smaller (yet correct) supports for each  $\mathbf{x}_k$ .
2. Lemma 1 and Lemma 2 should be thought of as the central results of this paper as they quantify the robustness of the support recovery in the noiseless case, the resulting ERC being merely direct consequences of the aforementioned lemmas. As mentioned previously, these lemmas can be used to produce theoretical analyses of SOMP for noisy scenarios while it is not the case for the ERC.

### 3 Sharpness of the bounds

In [6], it is shown that (ERC1) is sharp for OMP in the sense that it is possible to construct a measurement matrix  $\Phi_{\text{bad}}$  satisfying  $(1 - \delta_{|S|})/(\theta_{1,|S|}\sqrt{|S|}) = 1$  for which there exists a  $|S|$ -sparse signal  $\mathbf{x}_{\text{bad}}$  that OMP fails to recover. The sharpness property is immediately extended

to SOMP by noticing that if OMP fails to recover  $\mathbf{x}_{\text{bad}}$  on the basis of the measurement vector  $\mathbf{y}_{\text{bad}} = \Phi_{\text{bad}} \mathbf{x}_{\text{bad}}$ , then SOMP also fails with  $\mathbf{Y}_{\text{bad}} = \Phi_{\text{bad}} \mathbf{X}_{\text{bad}}$  where  $\mathbf{X}_{\text{bad}} = (\mathbf{x}_{\text{bad}}, \dots, \mathbf{x}_{\text{bad}})$  as both algorithms make the same decisions in this case.

Regarding (ERC2) and (ERC3), it has been shown in [18] that there exists a signal  $\mathbf{x}_{\text{bad}}$  of support  $S$  and a matrix  $\Phi_{\text{bad}}$  satisfying  $\delta_{|S|+1} = 1/\sqrt{|S|+1}$  for which OMP fails to recover the support of  $\mathbf{x}_{\text{bad}}$  on the basis of  $\mathbf{y}_{\text{bad}} = \Phi_{\text{bad}} \mathbf{x}_{\text{bad}}$ . Note that earlier works (see [19] and [25]) proved that the statement above holds with  $\Phi_{\text{bad}}$  satisfying  $\delta_{|S|+1} = 1/\sqrt{|S|}$ . Using an approach identical to that of the previous paragraph, one shows that this statement remains true for SOMP with  $\mathbf{Y}_{\text{bad}} = \Phi_{\text{bad}} \mathbf{X}_{\text{bad}}$  and  $\mathbf{X}_{\text{bad}} = (\mathbf{x}_{\text{bad}}, \dots, \mathbf{x}_{\text{bad}})$ . It shows that (ERC2) is near-optimal as, for  $|S| \rightarrow \infty$ , it boils down to the condition  $\delta_{|S|+1} < 1/\sqrt{|S|+1}$ . It can be shown that (ERC3) is also near-optimal but the discussion is more involved as  $\delta_{|S|}$  intervenes instead of  $\delta_{|S|+1}$ . In [23, Section 3], it is shown that  $\delta_{|S|+1} < 1/(|S| + 3 - \sqrt{2})$  implies (ERC3), therefore indicating that (ERC3) is also at least near-optimal.

## 4 Proofs

### 4.1 Proof of Lemma 1

The proof presented in this section is analog to what has been proposed in [7, 23, 24], the only difference being the additional quantities needed to deal with the MMV model.

The proof is decomposed in three steps:

1. Derive an upper bound on  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$  expressed as  $\theta_{1,|S|} \|\mathbf{z}^{(t)}\|_2$  where  $\mathbf{z}^{(t)}$  is to be specified in the detailed development.
2. Derive a lower bound on  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$  expressed as  $(1/\sqrt{|S|})(1 - \delta_{|S|}) \|\mathbf{z}^{(t)}\|_2$  where  $\mathbf{z}^{(t)}$  is identical for steps 1) and 2).
3. Compute the ratio of the lower bound to the upper bound and observe that the desired result is obtained due to the cancellation of the quantity  $\|\mathbf{z}^{(t)}\|_2$ .

Let us first tackle the quantity  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} = \max_{j \in \bar{S}} (\sum_{k=1}^K |\langle \mathbf{r}_k^{(t)}, \phi_j \rangle|)$  and define  $j^*(t) := \arg \max_{j \in \bar{S}} (\sum_{k=1}^K |\langle \mathbf{r}_k^{(t)}, \phi_j \rangle|)$ . Then, if  $c_k^{(t)} := \text{sign}(\langle \mathbf{r}_k^{(t)}, \phi_{j^*(t)} \rangle)$ , we have

$$\begin{aligned} \|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} &= \max_{j \in \bar{S}} \left( \sum_{k=1}^K |\langle \mathbf{r}_k^{(t)}, \phi_j \rangle| \right) \\ &= \left| \sum_{k=1}^K c_k^{(t)} \langle \mathbf{r}_k^{(t)}, \phi_{j^*(t)} \rangle \right| \\ &= \left| \left\langle \sum_{k=1}^K c_k^{(t)} \mathbf{r}_k^{(t)}, \phi_{j^*(t)} \right\rangle \right|. \end{aligned}$$

Since  $S_t \subseteq S$ ,  $\mathbf{r}_k^{(t)} = (\mathbf{I} - \mathbf{P}^{(t)}) \mathbf{r}_k$  belongs to  $\text{span}(\Phi_S)$  and can thus be expressed as a linear combination of the atoms whose indexes belong to  $S$  by means of  $\mathbf{r}_k^{(t)} = \Phi_S \mathbf{a}_k^{(t)}$  where  $\mathbf{a}_k^{(t)} \in \mathbb{R}^{|S|}$  contains the coefficients of the linear combination of interest. It is also worth defining the

extension  $\tilde{\mathbf{a}}_k^{(t)}$  of  $\mathbf{a}_k^{(t)}$  to  $\mathbb{R}^n$  by ensuring that  $\text{supp}(\tilde{\mathbf{a}}_k^{(t)}) \subseteq S$  and  $(\tilde{\mathbf{a}}_k^{(t)})_S = \mathbf{a}_k^{(t)}$ . Another relation of interest is  $\phi_{j^*(t)} = \Phi \mathbf{e}_{j^*(t)}$  where  $\mathbf{e}_{j^*(t)}$  denotes the  $j^*(t)$ th vector of the canonical basis of  $\mathbb{R}^n$ . Hence, using consecutively the equations of this paragraph and the definition of the ROC yields

$$\begin{aligned} \|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} &= \left| \left\langle \Phi \sum_{k=1}^K c_k^{(t)} \tilde{\mathbf{a}}_k^{(t)}, \Phi \mathbf{e}_{j^*(t)} \right\rangle \right| \\ &\leq \theta_{1,|S|} \left\| \sum_{k=1}^K c_k^{(t)} \tilde{\mathbf{a}}_k^{(t)} \right\|_2 \end{aligned}$$

where  $\|\mathbf{e}_{j^*(t)}\|_2$  is equal to 1. It is worth explicitly pointing out that the ROC definition is applicable in that case because the supports of  $\mathbf{e}_{j^*(t)}$  and  $\sum_{k=1}^K c_k^{(t)} \tilde{\mathbf{a}}_k^{(t)}$  are disjoint as  $j^*(t) \in \bar{S}$  and  $\text{supp}(\tilde{\mathbf{a}}_k^{(t)}) \subseteq S$  for  $1 \leq k \leq K$ .

The first step of the proof is now completed and the last problem to be dealt with is deriving a lower bound for  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$ . For any  $d_k^{(t)} \in \{-1; 1\}$ , we have  $|\langle \mathbf{r}_k^{(t)}, \phi_j \rangle| = |d_k^{(t)} \langle \mathbf{r}_k^{(t)}, \phi_j \rangle| = |\langle d_k^{(t)} \mathbf{r}_k^{(t)}, \phi_j \rangle|$ . In particular, it remains true for the choice  $d_k^{(t)} = c_k^{(t)}$ . Thus, by using the equation above and the triangle inequality, one obtains

$$\begin{aligned} \|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} &= \max_{j \in S} \left( \sum_{k=1}^K |\langle \mathbf{r}_k^{(t)}, \phi_j \rangle| \right) \\ &= \max_{j \in S} \left( \sum_{k=1}^K |\langle c_k^{(t)} \mathbf{r}_k^{(t)}, \phi_j \rangle| \right) \\ &\geq \max_{j \in S} \left| \left\langle \sum_{k=1}^K c_k^{(t)} \mathbf{r}_k^{(t)}, \phi_j \right\rangle \right| \\ &= \left\| \Phi_S^T \left( \sum_{k=1}^K c_k^{(t)} \mathbf{r}_k^{(t)} \right) \right\|_{\infty} \\ &\geq \frac{1}{\sqrt{|S|}} \left\| \Phi_S^T \left( \sum_{k=1}^K c_k^{(t)} \mathbf{r}_k^{(t)} \right) \right\|_2 \end{aligned}$$

where  $\Phi_S^T (\sum_{k=1}^K c_k^{(t)} \mathbf{r}_k^{(t)}) \in \mathbb{R}^{|S|}$ . Also, we have previously obtained  $\mathbf{r}_k^{(t)} = \Phi_S \mathbf{a}_k^{(t)}$ . The lower bound on  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$  is thus finally obtained by successively using the two previous relations and the inequality  $1 - \delta_{|S|} \leq \lambda_{\min}(\Phi_S^T \Phi_S)$  resulting from the RIP (see Section 1.3) in the following manner:

$$\begin{aligned} \|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty} &\geq \frac{1}{\sqrt{|S|}} \left\| \Phi_S^T \Phi_S \left( \sum_{k=1}^K c_k^{(t)} \mathbf{a}_k^{(t)} \right) \right\|_2 \\ &\geq \frac{1 - \delta_{|S|}}{\sqrt{|S|}} \left\| \sum_{k=1}^K c_k^{(t)} \mathbf{a}_k^{(t)} \right\|_2 \end{aligned}$$

where  $\|\sum_{k=1}^K c_k^{(t)} \mathbf{a}_k^{(t)}\|_2 = \|\sum_{k=1}^K c_k^{(t)} \tilde{\mathbf{a}}_k^{(t)}\|_2$ . The final result is now established by expressing the ratio of the lower bound on  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$  to the upper bound on  $\|\Phi_S^T \mathbf{R}^{(t)}\|_{\infty \rightarrow \infty}$ .  $\square$



## 4.2 Proof of Lemma 2

The proof consists in finding lower bounds on the ratio  $(1 - \delta_{|S|})/(\theta_{1,|S|}\sqrt{|S|})$  intervening in Lemma 1. For the first bound, it is sufficient to use the inequalities  $\delta_{|S|} \leq \delta_{|S|+1}$  and  $\theta_{1,|S|} \leq \delta_{|S|+1}$  [3, Lemma 2.1], for the numerator and the denominator respectively. The second bound is obtained by using the inequality  $\theta_{1,|S|} \leq \sqrt{|S|/(|S| - 1)}\delta_{|S|}$  on the denominator for  $|S| \geq 2$  [23, Section 2.3].  $\square$

## 4.3 Proof of Theorem 1

Let us first address the proof of (ERC1). At iteration 0, we have  $\mathbf{R}^{(t)} = \mathbf{Y}$  and Lemma 1 shows that a sufficient condition for SOMP to pick a correct atom is  $(1 - \delta_{|S|})/(\theta_{1,|S|}\sqrt{|S|}) > 1$  as it means that the highest metric is necessarily obtained for one of the correct atoms. Thus, at iteration 1, the condition  $S_1 \subseteq S$  is verified and Lemma 1 shows, once again, that a correct decision will be made. By repeatedly applying the same train of thought, one proves the theorem by induction. The remaining ERC are obtained in an identical manner by using the two bounds provided by Lemma 2 instead of that of Lemma 1.  $\square$

## Acknowledgments

The authors would like to thank the Belgian “Fonds de la recherche scientifique” for having funded this research.



## References

- [1] T. T. Cai and L. Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *Information Theory, IEEE Transactions on*, 57(7):4680–4688, 2011.
- [2] T. T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *Signal Processing, IEEE Transactions on*, 58(3):1300–1308, 2010.
- [3] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9):589–592, 2008.
- [4] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.
- [5] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.
- [6] W. Dan. A sharp rip condition for orthogonal matching pursuit. In *Abstract and Applied Analysis*, volume 2013. Hindawi Publishing Corporation, 2013.
- [7] W. Dan and R. Wang. Robustness of orthogonal matching pursuit under restricted isometry property. *Science China Mathematics*, 57(3):627–634, 2014.
- [8] M. Davenport, M. B. Wakin, et al. Analysis of orthogonal matching pursuit using the restricted isometry property. *Information Theory, IEEE Transactions on*, 56(9):4395–4401, 2010.
- [9] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive approximation*, 13(1):57–98, 1997.
- [10] J.-F. Determe, J. Louveau, L. Jacques, and F. Horlin. Simultaneous orthogonal matching pursuit with noise stabilization: theoretical analysis. *arXiv preprint arXiv:1506.05324*, 2015.
- [11] J. Ding, L. Chen, and Y. Gu. Robustness of orthogonal matching pursuit for multiple measurement vectors in noisy scenario. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 3813–3816. IEEE, 2012.
- [12] D. L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.
- [13] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.
- [14] Y. C. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *Information Theory, IEEE Transactions on*, 55(11):5302–5316, 2009.
- [15] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.
- [16] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier analysis and Applications*, 14(5-6):655–687, 2008.

- [17] E. Liu and V. N. Temlyakov. The orthogonal super greedy algorithm and applications in compressed sensing. *Information Theory, IEEE Transactions on*, 58(4):2040–2047, 2012.
- [18] Q. Mo. A sharp restricted isometry constant bound of orthogonal matching pursuit. *arXiv preprint arXiv:1501.01708*, 2015.
- [19] Q. Mo and Y. Shen. A remark on the restricted isometry property in orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 58(6):3654–3656, 2012.
- [20] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.
- [21] Y. Shen and S. Li. Sparse signals recovery from noisy measurements by orthogonal matching pursuit. *Inverse Problems and Imaging*, 9(1):231–238, 2015.
- [22] J. A. Tropp, A. C. Gilbert, and M. J. Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.
- [23] J. Wang, S. Kwon, and B. Shim. Near optimal bound of orthogonal matching pursuit using restricted isometric constant. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–7, 2012.
- [24] J. Wang and B. Shim. Improved recovery bounds of orthogonal matching pursuit using restricted isometry property. *arXiv preprint ArXiv:1211.4293*, 2012.
- [25] J. Wang and B. Shim. On the recovery limit of sparse signals using orthogonal matching pursuit. *Signal Processing, IEEE Transactions on*, 60(9):4973–4976, 2012.
- [26] R. Wu, W. Huang, and D.-R. Chen. The exact support recovery of sparse signals with noise via orthogonal matching pursuit. *Signal Processing Letters, IEEE*, 20(4):403–406, 2013.
- [27] W. Xu, Z. Li, Y. Tian, Y. Wang, and J. Lin. Perturbation analysis of simultaneous orthogonal matching pursuit. *Signal Processing*, 116:91–100, 2015.
- [28] M. Yang and F. de Hoog. Coherence and rip analysis for greedy algorithms in compressive sensing. *arXiv preprint arXiv:1307.1949*, 2013.
- [29] T. Zhang. Sparse recovery with orthogonal matching pursuit under rip. *Information Theory, IEEE Transactions on*, 57(9):6215–6221, 2011.